Course: INF2178

Student name: Jingwen Ji

Student number: 1006128101

Does the normalized whole brain volume of dementia patients change over the time of visits?

Introduction:

Dementia, a broad category of brain diseases characterized by a long-term and often gradual decrease in the ability to think and remember, significantly impacts a person's daily functioning. Among the various biomarkers used to understand and monitor the progression of dementia, the Normalized Whole Brain Volume (nWBV) stands out as a crucial measure. nWBV offers a quantitative glimpse into brain atrophy, a hallmark of many forms of dementia, including Alzheimer's disease.

Brain atrophy, or the loss of neurons and the connections between them, is a common feature of dementia. As dementia progresses, the brain visibly shrinks in size, a phenomenon that can be detected and measured through imaging techniques such as Magnetic Resonance Imaging (MRI). The nWBV is a metric derived from such images, representing the ratio of brain volume to the overall volume of the cranium. It provides a normalized measure that can be compared across individuals and over time within the same individual, offering insights into the rate and pattern of brain atrophy.

Understanding changes in nWBV is critical for several reasons. First, it helps in the early detection of cognitive impairment and dementia. Studies have shown that individuals with lower nWBV are at a higher risk of developing dementia, making it a potential predictive marker. Second, tracking changes in nWBV over time can offer valuable information about the progression of the disease, helping clinicians tailor interventions and support to the needs of the individual. Finally, nWBV serves as an important endpoint in clinical trials, providing objective evidence of the impact of therapeutic interventions on brain volume.

Research Question:

How does the progression of normalized whole brain volume (nWBV) differ among individuals diagnosed with dementia?

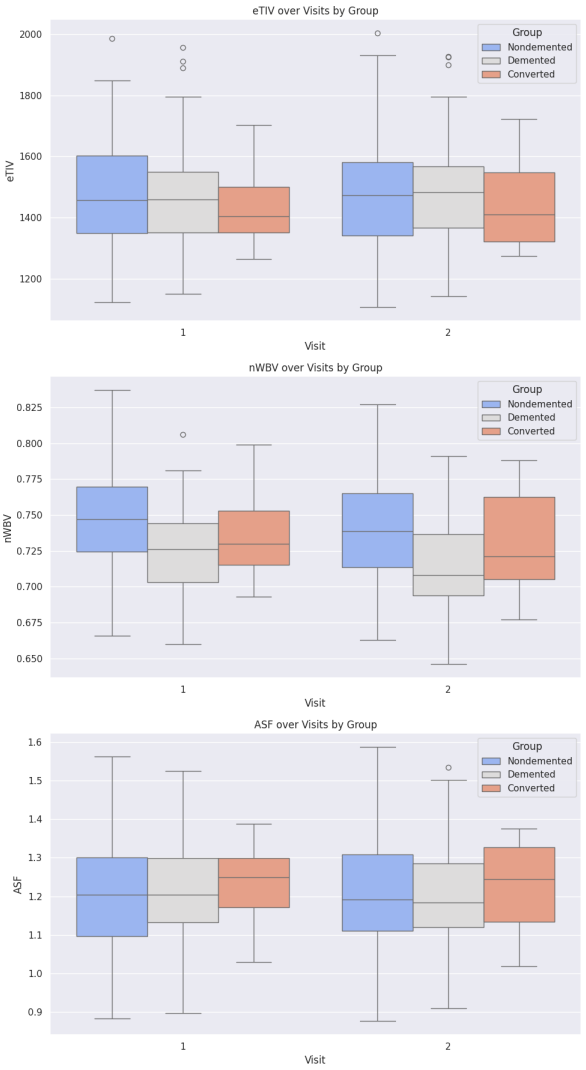
First of all, we display the columns of the dataset and their explanations.

Column	Description
Subject ID	A unique identifier assigned to each participant in the study, ensuring anonymity and facilitating longitudinal tracking.
MRI ID	A unique identifier for each MRI scan associated with a participant, allowing for the differentiation of multiple scans over time.
Group	Categorization of participants based on their diagnosis, typically into groups such as 'Nondemented', 'Demented', and 'Converted', to study the progression of cognitive impairment.

Column	Description
Visit	A numerical value indicating the sequence of visits for each participant, used to track longitudinal changes and the progression of dementia.
MR Delay	The time in days between the first visit and the subsequent visit(s), providing a timeline for longitudinal analysis.
M/F	The gender of the participant, indicated as 'M' for male and 'F' for female, considering the potential influence of gender on dementia progression.
SES	Socioeconomic status, often derived from occupation and educational background, potentially correlating with health outcomes and access to care.
MMSE	Mini-Mental State Examination score, a common measure of cognitive function used to screen for cognitive impairment and dementia.
CDR	Clinical Dementia Rating, a numerical scale used to quantify the severity of symptoms of dementia.
eTIV	Estimated Total Intracranial Volume, measuring the overall size of the brain, used in normalizing brain volume measurements.
nWBV	Normalized Whole Brain Volume, an indicator of brain atrophy relevant in the study of dementia and its progression.
ASF	Atlas Scaling Factor, used in the processing of MRI data to normalize brain images for comparison across participants.

Through simple data analysis, we find out the original data frame is pretty clean except there are some missing values in MMSE and SES columns. After imputing the missing values for MMSE and SES with their respective medians, there are no missing values left in the dataset across all columns. Imputing missing values with medians is a common strategy in data preprocessing, especially for numerical data, due to robustness to outliers, its applicability to ordinal data and its simplicity and computational efficiency.

Through the descriptive statistics, we could get a comprehensive overview of various measurements within the dataset. The average age of participants is approximately 76.4 years, ranges from 60 to 98 years among participants. The median SES score is 2, with a range from 1 to 5, indicating a middle to high socioeconomic status for most participants. The normalized whole brain volume has a mean of approximately 0.731, indicating the average proportion of brain volume relative to cranial volume. The MMSE scores range from 15 to 30, where a lower score indicates severe cognitive impairment. The average MMSE score is about 27.3 which suggests that, on average, participants have mild cognitive impairment.

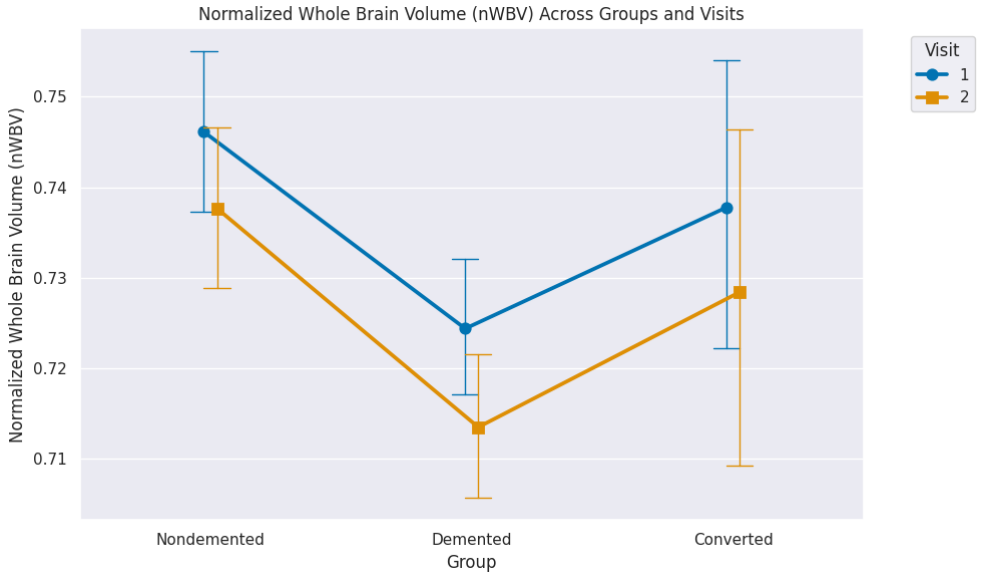


There is a noticeable decrease in the median nWBV from the first to the second visit across all groups, which may suggest progressive brain atrophy. The variability (IQR) of nWBV is quite similar across groups and visits, but the Nondemented group has a few outliers, particularly on the first visit. The decrease in nWBV could indicate an ongoing process related to dementia, while changes in ASF might relate to adjustments made during MRI processing across different time points or other factors.

Null Hypothesis (H0): There is no significant difference in the mean normalized whole brain volume (nWBV) between the different groups (Nondemented, Demented, Converted) over time.

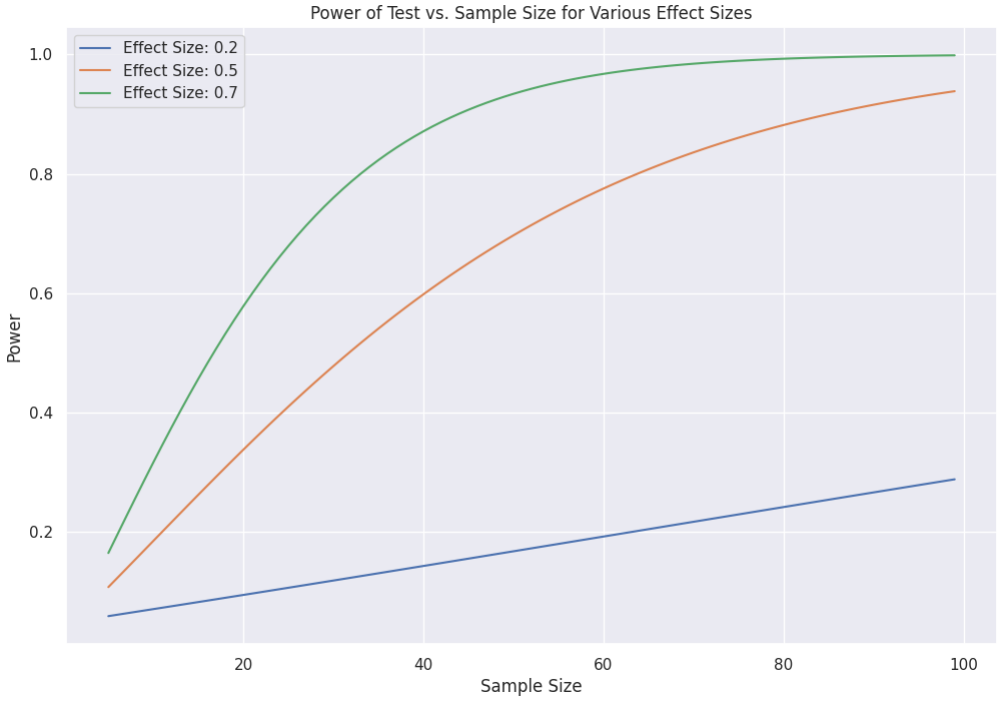
Alternative Hypothesis (H1): There is a significant difference in the mean nWBV between at least two of the groups over time.

Then we test the normality, homogeneity of variances, and the sphericity of the data. The normality tests performed using the Shapiro-Wilk test for each combination of 'Group' and 'Visit' within the dataset provide the p-values above the significance level of 0.05, suggesting that there is no significant evidence to reject the null hypothesis of normality. In other words, it appears that the nWBV data are normally distributed for each subgroup defined by 'Group' and 'Visit'. The results of Levene's test for homogeneity of variances (homoscedasticity) across different groups for each visit separately, p-values are 0.35422 and 0.742338 which well above the conventional threshold of 0.05, leading to the conclusion that variances of normalized whole brain volume (nWBV) are equal across the groups of Nondemented, Demented, and Converted. The condition of homogeneity of covariances is also satisfied as p-value is 1.0 indicate that there is no violation of the sphericity assumption in the data, meaning that the variances of the differences between the visits are equal.



For the **Nondemented group**, there is a notable decrease in mean nWBV from Visit 1 to Visit 2. In contrast, the **Demented group** shows an increase in mean nWBV from Visit 1 to Visit 2. The **Converted group** also shows an increase, but it starts with a lower mean nWBV at Visit 1 compared to the other two groups. We then perform post-hoc pairwise tests comparing the groups across different visits. **Visit 1 vs. Visit 2** : There is a highly significant difference in nWBV ($p < .001$), but the direction is not specified. Converted vs. Demented and Nondemented vs. Demented show a significant difference, indicated by the p-values of 0.181 and 0.0003 respectively. Converted vs. Nondemented does not show a statistically significant difference ($p = 0.526$). For Interaction effects (Visit * Group), there are no significant interaction effects between Visit and Group for any group comparisons at either visit, given the p-values are above the conventional threshold of 0.05.

The data suggests that while there are significant differences in nWBV across visits, the interaction effect between the group and visit is not significant, indicating that the changes in nWBV across visits do not differ significantly between groups.



The theoretical experiment with a **power=0.91**, **alpha=0.05** and **effect size=0.7** would need to have a sample size of 46 participants per group.